



Photometric Redshift Calibration Requirements for *WFIRST* Weak-lensing Cosmology: Predictions from CANDELS

Shoubaneh Hemmati¹ , Peter Capak¹ , Daniel Masters^{1,2} , Iary Davidzon¹ , Olivier Doré^{2,3} , Jeffrey Kruk⁴, Bahram Mobasher⁵, Jason Rhodes^{2,3} , Daniel Scolnic⁶, and Daniel Stern²

¹ IPAC, California Institute of Technology, 1200 East California Boulevard, Pasadena CA 91125, USA; shemmati@caltech.edu

² Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

³ California Institute of Technology, 1201 East California Boulevard, Pasadena, CA 91125, USA

⁴ NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

⁵ University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA

⁶ Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637, USA

Received 2018 August 30; revised 2019 April 14; accepted 2019 April 22; published 2019 June 3

Abstract

In order for the *Wide-Field Infrared Survey Telescope* (*WFIRST*) and other stage IV dark energy experiments (e.g., Large Synoptic Survey Telescope, LSST; and *Euclid*) to infer cosmological parameters not limited by systematic errors, accurate redshift measurements are needed. This accuracy can be met by using spectroscopic subsamples to calibrate the photometric redshifts for the full sample. In this work, we find the minimal number of spectra required for the *WFIRST* weak-lensing redshift calibration by employing the Self-Organizing Map (SOM) spectroscopic sampling technique. We use galaxies from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) to build the LSST+*WFIRST* lensing analog sample of $\sim 36,000$ objects and to train the LSST+*WFIRST* SOM. We find that 26% of the *WFIRST* lensing sample consists of sources fainter than the *Euclid* depth in the optical, 91% of which live in color cells already occupied by brighter galaxies. We demonstrate the similarity between faint and bright galaxies as well as the feasibility of redshift measurements at different brightness levels. Our results suggest that the spectroscopic sample acquired for calibration to the *Euclid* depth is sufficient for calibrating the majority of the *WFIRST* color space. For the spectroscopic sample to fully represent the synthetic color space of *WFIRST*, we recommend obtaining additional spectroscopy of ~ 0.2 – $1.2k$ new sources in cells occupied by mostly faint galaxies. We argue that either the small area of the CANDELS fields and the small overall sample size or the large photometric errors might be the reason for no/fewer bright galaxies mapped to these cells. Acquiring the spectra of these sources will confirm the above findings and will enable the comprehensive calibration of the *WFIRST* color–redshift relation.

Key words: dark energy – galaxies: distances and redshifts

1. Introduction

Revealing the nature of the dark energy driving cosmic acceleration and testing general relativity on cosmological scales are essential pieces to complete our understanding of modern cosmology and physics. To achieve these goals, the next-generation large cosmology surveys will make precision measurements of the expansion history of the universe as well as the growth rate of large-scale structures (LSSs) using various techniques (Spergel 2015).

Samples of Type Ia supernovae (SN Ia) constrain cosmological parameters (e.g., the expansion rate of the universe) by providing measurements of cosmological distances as a function of redshift (e.g., Riess et al. 1998; Perlmutter et al. 1999). Complementary distance scale measurements can be obtained from baryon acoustic oscillation (BAO) imprints in the power spectrum of the cosmic microwave background and in the LSSs of galaxies at lower redshifts (e.g., Zhan & Knox 2006; Benítez et al. 2009; Aubourg et al. 2015). Weak gravitational lensing of distant galaxies by the gravitational field of matter inhomogeneities in the LSS, or cosmic shear, provides another powerful tool for constraining the power

spectrum of dark and luminous matter in the universe (e.g., Blandford et al. 1991; Blandford & Narayan 1992). Weak-lensing cosmology requires both redshift estimates and shape measurements of statistical samples of galaxies (e.g., Hu 1999).

Upcoming stage IV dark energy experiments aimed for the 2020s (see Albrecht et al. 2006) will improve current measures of the distance and cosmic expansion history (with uncertainties $\sim 1\%$ – 3%) as well as matter clustering (with uncertainties $\sim 5\%$ – 10%) to 0.1% – 0.5% precision, while also extending them to previously unexplored redshift regimes. Careful calibration is required such that the cosmological inferences will not be limited by systematic errors (Spergel 2015).

Accurate redshifts are needed for all three techniques mentioned above (SN Ia, LSS BAO, and weak-lensing tomography). While SN Ia and BAO studies usually employ a spectroscopic sample, obtaining spectroscopic redshifts for hundreds of millions to billions of faint galaxies needed for weak-lensing analysis is not practical. Therefore, highly accurate photometric redshifts, trained and validated using a training sample of spectroscopic data, are required.

Several recent studies (e.g., Cunha et al. 2012; Masters et al. 2015; Newman et al. 2015) have investigated the best spectroscopic sampling strategy in order to train higher quality, lower scatter photo- z with less systematic errors for different cosmological surveys. Carrasco Kind & Brunner (2013) showed that a random selection of galaxies to create a



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

spectroscopic training sample is not optimal. Recent work has suggested spatial cross-correlation-based techniques relating the photometric redshifts with a reference spectroscopic sample as a solution (e.g., Newman 2008; Newman et al. 2015). These techniques also require a large spectroscopic sample. However, their main advantage is that the spectroscopic sample does not need to be representative of different galaxy types (i.e., bright emission-line galaxies, which are easy spectroscopic targets, can be used for calibration).

A completely data-driven technique of selecting optimal spectroscopic samples to meet the cosmological requirements was introduced by Masters et al. (2015, hereafter M15). This technique uses a machine-learning algorithm called the Self-Organizing Map (SOM; Kohonen 1982) to reduce the multi-dimensional color space of galaxies defined by a photometric survey to two dimensions (hence maps). This empirical color-mapping method allows us to focus our spectroscopic efforts on undersampled regions of the galaxy parameter space. M15 explored different SOM-based targeting strategies and estimated the required spectroscopy for the *Euclid* mission (Laureijs et al. 2011). This approach is the basis of a large, ongoing spectroscopic program, the Complete Calibration of the Color–Redshift Relation (C3R2) survey, designed to calibrate the color–redshift relation to the *Euclid* depth (Masters et al. 2017, 2019). Recently, Sánchez & Bernstein (2019) presented a framework based on a hierarchical Bayesian model to infer redshift distributions from the combination of galaxy colors and clustering information. SOMs can also be used to define galaxy phenotypes based on these Bayesian schemes, specifically where more limited and noisy colors exist for galaxies.

In this paper, as part of the High Latitude Survey (HLS; Doré et al. 2018) science investigation team, we extend the previous analysis of M15 to estimate the additional spectroscopic sample required to meet the *Wide-Field Infrared Survey Telescope* (*WFIRST*) cosmological requirements. *WFIRST* is a NASA flagship mission using a 2.4 m telescope to provide measurements of the expansion history of the universe and growth of structure to better than 1% (Spergel 2015). For weak-lensing analysis, the *WFIRST* HLS is currently planning to image 2227 deg^2 in four near-infrared (near-IR) bands (*Y*, *J*, *H*, and *F184*) spanning the range from $0.92\text{--}2.00 \mu\text{m}$ to magnitudes $25.8\text{--}26.7$ (depending on band), significantly fainter compared to the near-IR depths in the *Euclid* survey (~ 24.5). The near-IR filters alone are not sufficient for precise photo-*z* estimation. Multiband optical observations need to be combined with *WFIRST* to fulfill the redshift requirements. Such observations will be available through the Hyper Suprime-Cam on Subaru or by the Large Synoptic Survey Telescope (LSST).

In Section 2, we simulate a data-driven photometric catalog using deep observations from the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) to replicate the LSST+*WFIRST* lensing sample. In Section 3, we briefly review the SOM technique, train the SOM with the LSST+*WFIRST* analog catalog, and test for its accuracy in representing the data. We check for the effects of cosmic variance in Section 4 and address the additional spectroscopy needed to meet *WFIRST* cosmology requirement in Section 5. Section 6 summarizes the results of this work and discusses sources of uncertainty. Throughout this paper, all magnitudes are expressed in the AB system (Oke & Gunn 1983), and we use the concordance

Λ CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

2. *WFIRST* + LSST Weak-lensing Sample

2.1. CANDELS Galaxy Sample

CANDELS photometric catalogs are the optimum choice for testing *WFIRST* HLS redshift requirements as they are *H*-band-selected and provide deeper multiband observations compared to *WFIRST*. CANDELS obtained very deep near-IR observations of five different fields using the *Hubble Space Telescope* (*HST*). These observations are homogeneously combined with the wealth of ancillary space- and ground-based data available from UV to X-ray.

Source detection in all CANDELS catalogs was conducted in the Wide Field Camera 3 (WFC3) F160W band (at $1.6 \mu\text{m}$), and photometry is generated using the Template FITting algorithm (Laidler et al. 2007). Details of CANDELS observations and photometric catalogs for each of the five fields used in this study can be found in the published catalog paper for each field: the GOODS South field (Guo et al. 2013), the UDS field (Galametz et al. 2013), the COSMOS field (Nayyeri et al. 2017), the Extended Groth Strip field (Stefanon et al. 2017), and the GOODS North field (G. Barro et al. 2019, in preparation). The photometric redshifts published by the CANDELS team, which we use in this work are based on combining results from multiple teams, each using a different combination of photometric redshift code, library of template spectral energy distributions (SEDs), and priors (see, e.g., Dahlen et al. 2013). These measurements have an rms of ~ 0.03 with an outlier fraction of at most $\sim 3\%$, measured from a spectroscopic sample. Dahlen et al. (2013) found a strong magnitude dependence in the accuracy of photometric redshifts, and hence, the reported uncertainties might be slightly underestimated for spectroscopic samples biased toward brighter objects. In this work, where we use a subsample of CANDELS galaxies suitable for lensing analysis (discussed in Section 2.3), the depth of the spectroscopic sample is comparable to our targets, and hence, the level of uncertainty is still below what is expected for individual photo-*z* precision for weak-lensing analysis in stage IV cosmology experiments (e.g., Bordoloi et al. 2010).

The total area covered by the CANDELS (wide) observations is $\sim 0.2 \text{ deg}^2$, reaching a 5σ limiting magnitude depth of ~ 26.5 in the WFC3 F160W observations, more than a magnitude deeper than the planned depth for *WFIRST* (see Figure 1). The small area covered by the CANDELS fields compared to the *WFIRST* HLS can raise concerns about the effects of cosmic variance in the analysis. We address this issue briefly in Section 4.

2.2. From CANDELS to LSST + *WFIRST* Photometry

To estimate the photometry of CANDELS galaxies in each of the nine LSST and *WFIRST* filters, we use a linear interpolation of the photometry in the deepest CANDELS filters that straddle the LSST and *WFIRST* filters in each field. Figure 1 shows the LSST and *WFIRST* filters and their expected 5σ limiting depths (HLS-expected depths are shown for *WFIRST*) in comparison to the CANDELS GOODS-S filters used to estimate the photometry in the new filters. Table 1 lists the CANDELS filters used in each field for these measurements. Figure 2 shows five sample SEDs and the

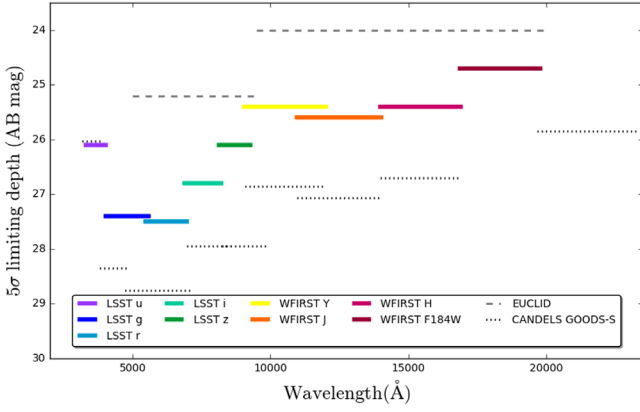


Figure 1. Expected 5σ limiting magnitudes of the LSST and the *WFIRST* high latitude survey (HLS) filters for galaxies ($r_{1/2} = 0''.3$) plotted as solid colored lines. We use deeper or similar depth photometry in the same wavelength range as probed by LSST+*WFIRST* from CANDELS catalogs in five fields to estimate LSST+*WFIRST* photometry. 5σ limiting AB magnitude of CANDELS GOODS-S filters for galaxies ($r_{1/2} = 0''.3$) are plotted with dotted black lines (CANDELS filters used are listed in Table 1). *Euclid*'s expected r_z as well as NIR 5σ depths are also overplotted as dashed gray lines for comparison.

estimated LSST+*WFIRST* photometry from linear interpolation. We extensively tested this method of linear interpolation to estimate the broadband photometry. The interpolation technique reproduces more realistic color distributions compared to fitting galaxies with theoretical (or empirical) model SEDs and convolving best-fitted model SEDs with the broadband filters. This technique is model independent and data-driven, and therefore lacks SED-modeling uncertainties. Furthermore, the CANDELS bands are well matched to the *WFIRST*/LSST bands and deeper, so the linear interpolation introduces minimal errors with respect to the true photometry.

2.3. *WFIRST* Lensing Analog Sample

As the main purpose of this study is to find the spectroscopic sample requirement for weak lensing, we apply a cut to the photometric catalog to exclude galaxies that will not benefit the lensing analysis (C. Hirata 2018, private communication). This cut removes galaxies with shape distortion measurements that will not be accurate enough for weak-lensing analysis. Figure 3 shows this lensing criterion applied to the sample based on the galaxy FWHM in the F160W filter and the average $J+H$ magnitude. This criterion derives from requiring the galaxy to be resolved, the average $J+H$ flux having signal-to-noise (S/N) > 18 , and the ellipticity uncertainty < 0.2 . The limiting flux depends on the galaxy size. The cut applied is deliberately inclusive, and if specific regions of the color–magnitude–size space are found to be difficult for lensing analysis and shape measurement, those could be excluded from the lensing analysis later. The lensing cut applied here also excludes the very faint galaxies that exist in deep CANDELS data but will not be in the *WFIRST*, due to magnitude limits.

The final photometric catalog of the lensing sample consists of 36,612 galaxies with estimates of their photometry in LSST optical and *WFIRST* near-IR filters, as well as their photometric and spectroscopic (when available) redshifts, and physical sizes. Figure 4 shows the distribution of this sample in the H band and redshift, with the majority of sources at $H_{\text{WFIRST}} \sim 24$ mag and $z \sim 1$.

3. 2D Map of the *WFIRST* Color Space

SOMs offer an optimized and efficient way to target sources for spectroscopy from the less explored regions of galaxy color space and provide the means to calibrate the photometric redshifts using the acquired spectroscopic sample. In short, the SOMs introduced in the 1980s by Kohonen are a class of unsupervised artificial neural networks that reduce the dimensionality of a multi-dimensional parameter space (the color space of galaxies in this case), while preserving the topology in the parameter space. In other words, similar objects in multidimensional parameter space remain neighboring on two-dimensional grids (maps). Therefore, SOMs are also an optimized way of visualizing multidimensional parameter space.

3.1. Training the SOM with *WFIRST*-analog Data

In this work, we use the SOMPY package, a Python library for SOMs. We assume a rectangular topology and 80×60 cells. This grid size is optimized to represent the color space of the training sample (in Section 5, we discuss the effects of choosing different grid sizes). Similar to most other artificial neural networks, there is a training phase and a mapping phase. We use the LSST+*WFIRST* colors of galaxies in the lensing sample ($u_{\text{LSST}} - g_{\text{LSST}}, g_{\text{LSST}} - r_{\text{LSST}}, r_{\text{LSST}} - i_{\text{LSST}}, i_{\text{LSST}} - z_{\text{LSST}}, z_{\text{LSST}} - Y_{\text{WFIRST}}, Y_{\text{WFIRST}} - J_{\text{WFIRST}}, J_{\text{WFIRST}} - H_{\text{WFIRST}},$ and $H_{\text{WFIRST}} - F184_{\text{WFIRST}}$) to train the SOM. Each cell in the SOM will have a weight vector with the dimensions of the input training data (eight dimensions here). We assign the initial weights by principal component analysis. During the competitive training phase, weight vectors adapt themselves to represent the distribution and topology of the input data. This is done by finding the best matching unit (BMU) of the SOM for each data point in the training set and bringing the weight vector of the BMU, and its neighbors, closer to the training data point. The magnitude and radius of the change decrease over time until the SOM represents the data well. SOMPY does the training in batch mode, where the SOM is exposed to the entire training sample in each epoch. Batch-mode training is generally expected to be quicker and result in a more stable network. Figure 5 represents our trained SOM colored by each component (color) of the final weight vector at each cell. Each position on the two-dimensional map points to a position in the eight-dimensional color space.

The final quantization error of the SOM, which is defined as the mean of the Euclidean distances of all training data to their BMUs, is 0.81. To further verify that the SOM is trained properly and that the weight vectors of the neurons (cells) adequately represent the training data, we compare the distribution of SOM cell colors to lensing sample colors in Figure 6. The identical range and shapes of the distributions confirm qualitatively that the SOM cells represent the training data.

3.2. Mapping *WFIRST* Colors to Redshift

Figure 7 presents the redshift information of the lensing galaxy sample on the SOM. We map galaxies in the LSST+*WFIRST* analog catalog back to the trained SOM and color-code the SOM by the median photometric redshifts of galaxies in each cell (top left panel), where the photometric redshifts are measured by the CANDELS team (Dahlen et al. 2013; Nayyeri et al. 2017; Stefanon et al. 2017; G. Barro et al. 2019, in preparation). The smoothness of the photometric

Table 1
CANDELS Filters Used in Each Field to Create the LSST+*WFIRST* Catalog

Field	Filters ^a										
GOODS-S	U_{VIMOS}	F435W	F606W	F775W	F814W	F850lp	F098W	F105W	F125W	F160W	$K_{\text{SHAWK-I}}$
GOODS-N	U_{KPNO}	F435W	F606W	F775W	F814W	F850lp	F105W	F125W	F160W		K_{SCFHT}
EGS	U_{CFHT}	g_{CFHT}	F606W	r_{CFHT}	i_{CFHT}	F814W	z_{CFHT}	F125W	F160W		K_{SCFHT}
UDS	U_{CFHT}	B_{subaru}	F606W	R_{subaru}	i_{subaru}	F814W	z_{subaru}	$Y_{\text{HAWK-I}}$	F125W	F160W	$K_{\text{SHAWK-I}}$
COSMOS	U_{CFHT}	B_{subaru}	F606W	r_{subaru}	i_{CFHT}	F814W	z_{CFHT}	Y_{UVISTA}	F125W	F160W	K_{UVISTA}

Note.

^a Refer to CANDELS catalog papers a for detailed description of observations in each filter.

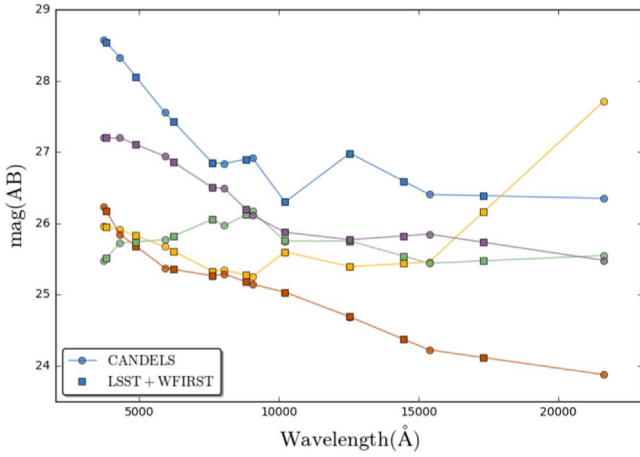


Figure 2. LSST- and *WFIRST*-measured photometry for five sample galaxies from CANDELS GOODS-S (squares). Solid lines and circles show the CANDELS broadband filters used for the linear interpolation. Each color represents a sample galaxy.

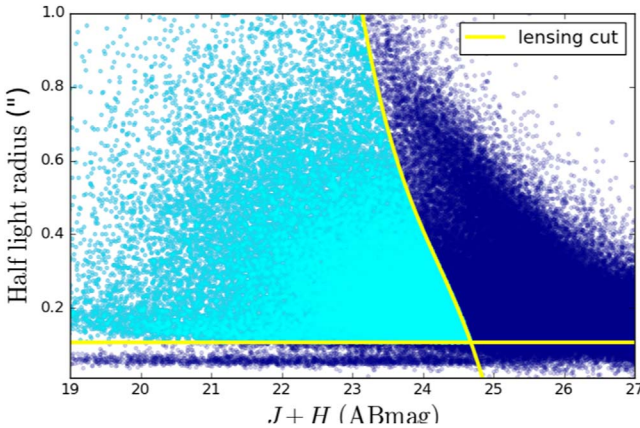


Figure 3. Subsample required for lensing analysis (aqua dots) selected from the LSST+*WFIRST* simulated photometric catalog (dark blue dots) by a cut based on size and $J+H$ AB magnitude (yellow solid lines), which requires the galaxy to be resolved, have a high S/N (>18), and a small ellipticity uncertainty (<0.2).

redshift map indicates that the combined colors in the LSST and *WFIRST* filter sets are adequate for photometric redshift estimation. A redshift uncertainty map defined as $\frac{\sigma_z}{1 + \text{photo}z_{\text{median}}}$,

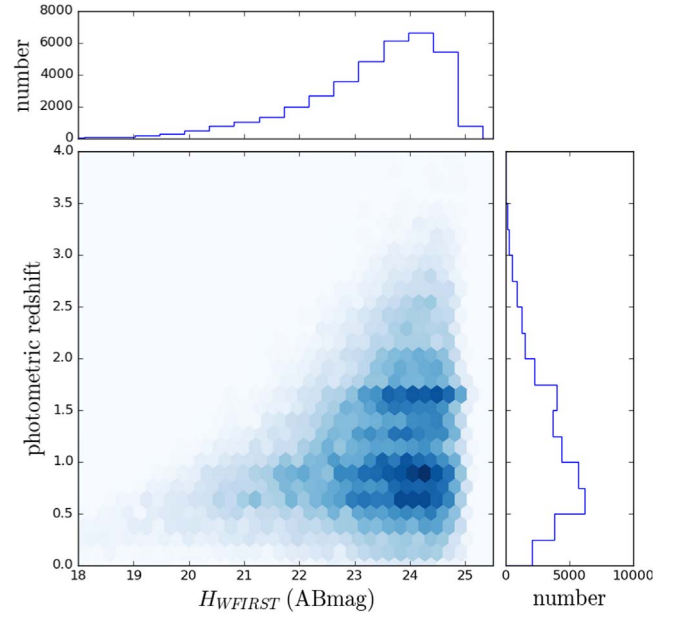


Figure 4. Density of the simulated lensing sample (a total of 36,612 galaxies selected from all five CANDELS fields that meet the lensing cut criteria) in the photometric redshift vs. H_{WFIRST} magnitude plane plotted as blue hexagons with darker colors representing higher densities of objects.

where σ_z is the standard deviation of the redshifts of galaxies in each cell, is shown in the bottom left panel of Figure 7. The average uncertainty per cell is small, of the order of 0.04, while larger uncertainties (~ 0.2) are found on the boundaries between high- and low-redshift regions of the SOM. High-confidence spectroscopic redshifts in CANDELS, where available, are also shown on the SOM (top right panel) and cover $\sim 57\%$ of the color space. These are from public spectroscopic redshifts available in the CANDELS fields, i.e., the CANDELS public compilation of spectroscopic redshifts (by N. Hathi), the MOSDEF public spectroscopic redshift catalog (Kriek et al. 2015), and the 3D-HST (Brammer et al. 2012) catalog of grism spectroscopy. While more than half of the SOM cells are occupied by at least one high-quality spectroscopic redshift, this corresponds to only $\sim 20\%$ of the sample (i.e., 7453 spectra). Visual comparison with the photometric redshift map shows good agreement between the photometric and spectroscopic redshifts. As discussed thoroughly in M15, using the SOM technique in calibrating the redshifts has the advantage of showing whether the training

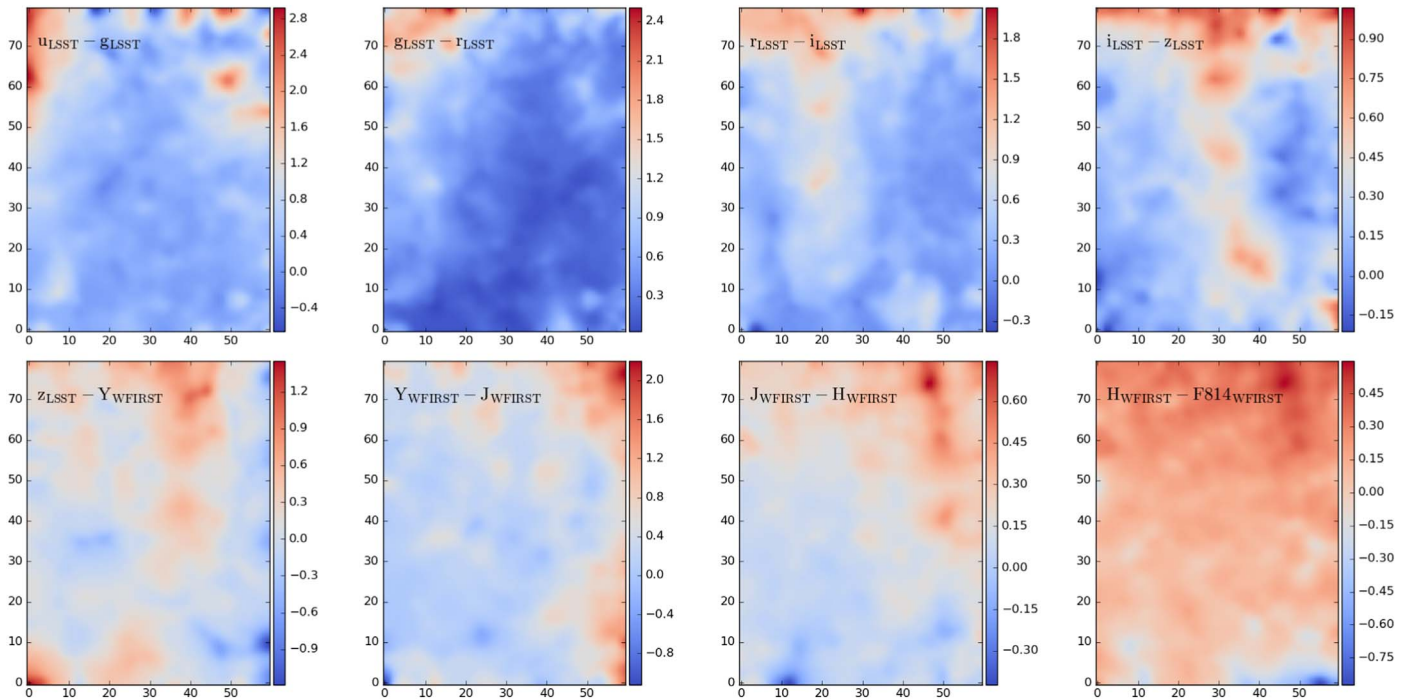


Figure 5. LSST and *WFIRST* colors of the trained SOM at each cell from top left to bottom right, color-coded by $u_{\text{LSST}} - g_{\text{LSST}}$, $g_{\text{LSST}} - r_{\text{LSST}}$, $r_{\text{LSST}} - i_{\text{LSST}}$, $i_{\text{LSST}} - z_{\text{LSST}}$, $z_{\text{LSST}} - Y_{\text{WFIRST}}$, $Y_{\text{WFIRST}} - J_{\text{WFIRST}}$, $J_{\text{WFIRST}} - H_{\text{WFIRST}}$, and $H_{\text{WFIRST}} - F814_{\text{WFIRST}}$. SOM is selected to be a mesh of 80×60 cells. The axes are arbitrary and each position on the two-dimensional map points to a position in the eight-dimensional color space.

sample is representative of the color distribution of galaxies in a survey. The high-confidence spectroscopic redshifts in each cell are necessary to calibrate the redshift map and hence the color–redshift relation to the accuracy needed by cosmology. The redshift bias parameter is also estimated (bottom right panel) and is defined as $\left(\frac{|\Delta z|_{\text{median}}}{1 + \text{spec}z_{\text{median}}} \right)$, where $|\Delta z|_{\text{median}} \equiv \text{median}(\text{photo}z - \text{spec}z_{\text{median}})$ and has lower values compared to the photometric uncertainty map with a median of ~ 0.03 . Most of the higher uncertainty cells seen in the photometric redshift uncertainty map (e.g., lower right and left corners of Figure 7) coincide with those where high-confidence spectroscopic data are missing. This is due to biases in spectroscopic samples, not covering all of color space, and larger photometric uncertainties leading to less certain photometric redshifts in these regions.

3.3. Beyond Redshifts and Broadband Photometry

Thus far we have trained and tested the LSST+*WFIRST* SOM using the eight optical and near-IR colors of the lensing analog sample. In this section, we go one step further by analyzing the high-resolution spectra of a sample of galaxies across the SOM. We draw galaxies from the VIMOS VLT Deep Survey (VVDS), which is a comprehensive deep galaxy spectroscopic redshift survey conducted by the VIMOS collaboration with the VIMOS multislit spectrograph at the ESO-VLT (Le Fèvre et al. 2013). The VVDS spectra also span a broad range of magnitudes with no cut on object properties other than their *I*-band magnitude. As a result, we can test how rapidly spectral properties change with position on the SOM.

Figure 8 shows some of the VVDS $z \sim 1$ galaxies with high-confidence spectroscopic redshifts that grouped in two distant regions of the SOM. These galaxies lie in the $z \sim 1$ region, as shown from the underlying color of the SOM (left panel). We

smooth and normalize the 1D spectra of these sources and shift them to $z = 0$ to compare. Spectra of galaxies mapped to upper and lower regions of the SOM are shown in the right panel of Figure 8. This figure shows that the SOM, trained by broadband colors, can also statistically group galaxies with similar spectral features closer together. This is akin to previous works using broadband photometry to estimate emission-line ratios in galaxies (e.g., Faisst et al. 2016). Here, for instance, the galaxies in the lower parts of the SOM show strong nebular emission lines (e.g., [O II]) while the upper ones do not. This test illustrates how the SOM classifies galaxies by their intrinsic spectral shape, and that high-resolution spectral features are captured by the SOM to some extent. The amount of high-resolution information is clearly limited by the photometric precision and spectral resolution of the filter set, but is sufficient to capture some key features.

The detailed quantification of spectral similarity is not straightforward and is beyond the scope of the current paper. Here we present a simple χ^2 analysis and leave a more detailed exploration of systematics and recovery of spectral features with simulated spectra to a future paper (S. Hemmati et al. 2019, in preparation). The χ^2 difference between the spectra shown in the right panel of Figure 8 is of the order ~ 0.1 and ~ 0.15 among spectra in the top and bottom panels, respectively. The χ^2 increases to more than 1 when spectra from the top and bottom panels are compared. In Figure 9, we map all 460 galaxies with high-quality spectroscopic redshifts from VVDS to the SOM (redshift range $0 < z < 3.5$). Then, we randomly pick 2000 pairs of galaxies, where the second galaxy had to be within a $dz = 0.1$ from the first random selection. We measured their distance on the SOM and the χ^2 difference between their smoothed spectra. The overall trend of decreasing similarity between spectra with increasing distance on the SOM can be clearly seen from this exercise. However, we note that while χ^2 does present a measure of similarity, it is

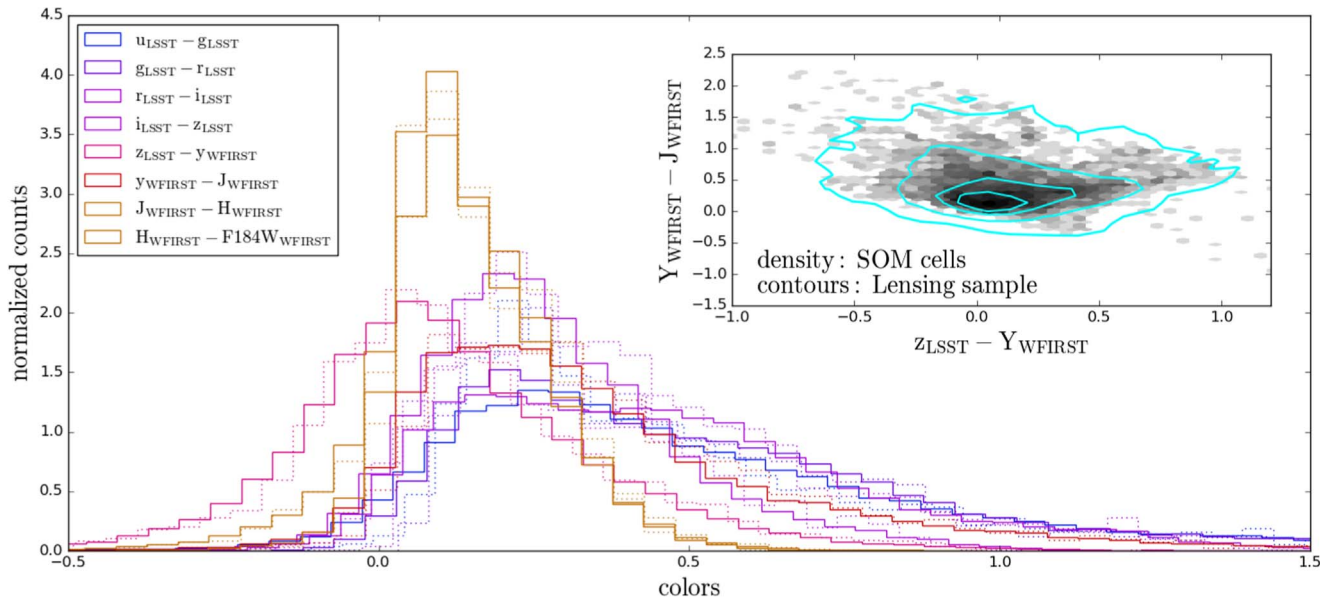


Figure 6. Comparison of normalized distribution of colors of galaxies in the input lensing catalog (solid) and colors of SOM cells (dashed) in the eight colors used for training the SOM. The inset shows an example color–color distribution ($Y_{WFIRST} - J_{WFIRST}$ vs. $Z_{LSST} - Y_{WFIRST}$) of the SOM cells with the gray density plot, and the aqua contours represent the lensing sample distribution with 10, 100, 500, and 1000 levels. The similarity between the distributions proves that the SOM is trained well and that the SOM cells represent the training data accurately.

not the best way to distinguish similar features such as emission lines between noisy spectra. Also, while distance on the SOM is an overall measure of how different shapes of galaxy SEDs are, a more careful analysis would be to use derivative-based clustering on the SOM, as distances on SOMs are not preserved after projection from multidimension.

4. Cosmic Variance

The CANDELS survey area is much smaller than what would be covered by the *WFIRST* HLS, so it is important to determine the effects of cosmic variance on redshift calibration. Despite this concern, the depth of multiband observations by CANDELS makes it the most favorable option to simulate the photometry of *WFIRST* galaxies. The effect of cosmic variance on measurement error has been claimed to be an order of magnitude larger than Poisson errors for determining redshift distributions (e.g., Newman & Davis 2002; Cunha et al. 2012). Newman et al. (2015) suggested that a minimum of 10–20 fields of $\sim 20'$ diameter each (to cover multiple correlation lengths) is needed in order to meet the requirements for photometric redshift calibration. Here, the CANDELS fields cover only five widely separated fields in the sky, covering a total of $\sim 0.2 \text{ deg}^2$.

We only briefly explore the effect of the small area in this work and defer the full quantitative analysis to a future paper. We visually compare the SOM trained by our *WFIRST* lensing sample to the *Euclid* SOM (Masters et al. 2017), which is trained using the combination of the COSMOS (Scoville et al. 2007), SXDS (Furusawa et al. 2008), and VVDS (Le Fèvre et al. 2005) surveys covering 3.8 deg^2 and is therefore 19 times larger. Figure 10 presents this comparison on three color–color plots with the density of colors of the *WFIRST* lensing sample SOM shown with shades of blue and the *Euclid* SOM overplotted using aqua contours. The range and shapes of these distributions are visually identical. This indicates that increasing the area ~ 20 times does not expand the color space,

but rather fills the gaps between the already covered color space.

Photometric redshifts measured from multiband SEDs need spectroscopic observations to calibrate the systematics in redshift measurements, due to the lack of precision in low-resolution SEDs. Selecting spectroscopic samples of galaxies randomly or based on environment for redshift calibration suffers from cosmic variance, as by definition they do not cover the whole range of possible SED shapes. Selecting the spectroscopic calibration sample systematically from the well-occupied color space instead should not suffer as much from the loss of different types of galaxies. However, we note that the quantification of cosmic variance in the redshift calibration with SOMs for weak lensing is critical and cannot be done with these small observed data sets. P. Capak et al. (2019, in preparation) explored the effect of cosmic variance in more depth, exploiting large cosmological simulations of galaxies, comparable in size to that of the *WFIRST* HLS.

5. Optimal Sampling Technique to Meet Weak-lensing Redshift Requirements

Different spectroscopic sampling strategies from the SOM were explored in M15 to calibrate the $\langle z \rangle$ of the tomographic redshift bins to the required level for weak-lensing cosmology ($\Delta \langle z \rangle < 0.002(1 + \langle z \rangle)$; see, e.g., Kitching et al. 2008; Bordoloi et al. 2010). M15 showed that with simplifying assumptions, if $\sigma_{\langle z \rangle}$ in the SOM cells are of the order of ~ 0.05 , with ~ 600 color cells (c) in each tomographic bin and assuming that the mean redshift of each cell is known (~ 1 spectrum per cell), the calibration requirement can be met (see M15 for details; in short, $\Delta \langle z \rangle \simeq \sigma_{\langle z \rangle} / \sqrt{c}$). In M15, this meant a total of $\sim 10,000$ – $15,000$ spectra, much lower compared to estimates of direct calibration through random sampling. The gain in statistical precision from the SOM method compared to direct sampling is attributed to the systematic way the full color space is sampled.

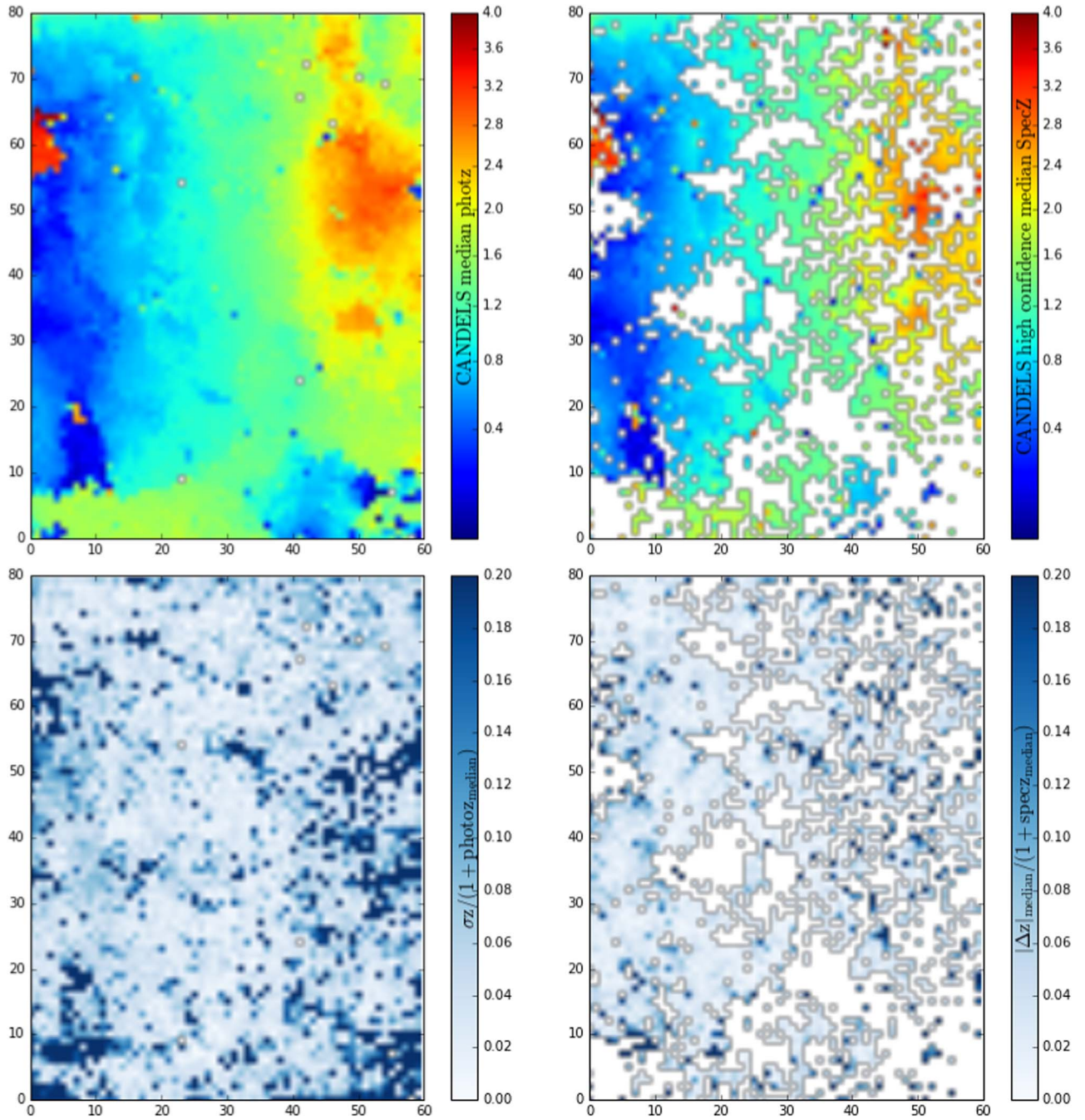


Figure 7. Distribution of the median photometric redshifts (top left), median of the high-confidence spectroscopic redshifts (top right), the full scatter σ_F of the photometric redshifts defined as $\sigma_{\text{photoz}}/(1 + z_{\text{median}})$ (bottom left), and the redshift bias, defined as $|\text{median}[\text{specz} - \text{photoz}|/(1 + \text{specz})$ (bottom right) on the SOM.

The average redshift uncertainties in the LSST+*WFIRST* SOM provided here are smaller compared to the estimates in M15 (~ 0.03 versus ~ 0.05), due to the higher S/N in *WFIRST* observations. The average number of cells needed in each tomographic bin to meet the calibration requirement (based on the simplified assumptions mentioned above) is then reduced to ~ 200 . Given an 80×60 grid for the SOM, there can be 24 tomographic bins of this size, and to fill the SOM with at least one spectroscopic redshift per cell, ~ 5000 spectra would be needed. This does not correspond to 5000 new spectra, as many CANDELS galaxies have prior spectroscopic observations (see Figure 7). Also, there are galaxies with almost identical SED

shapes in other fields of the sky, which we can use to calibrate this SOM (discussed in the following subsections).

It is important to note that the number of spectra needed for calibration is independent of the initial size of the rectangular grid chosen for the SOM. Having a smaller rectangular grid will increase the average redshift scatter in SOM cells. This is because more distant SEDs are grouped together in a cell as compared to having a larger grid, which leads to having a higher redshift scatter in each cell, and therefore more than one spectra per cell is needed to calibrate the photometric redshifts. On the other hand, increasing the grid size will lead to a smaller average redshift scatter in a cell. For instance, the average

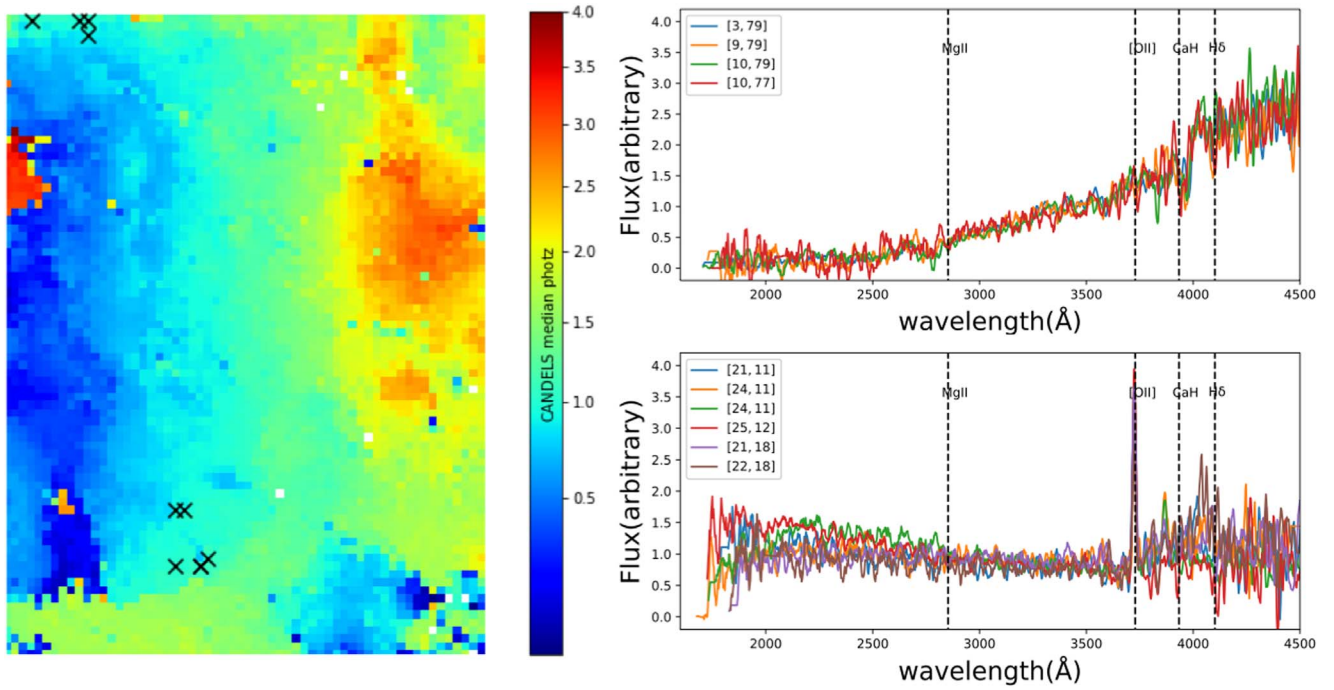


Figure 8. Ten $z \sim 1$ galaxies from the VVDS that have high-quality 1D spectra are mapped to the LSST+*WFIRST* SOM. In the left panel, we show where these 10 galaxies sit on the LSST+*WFIRST* SOM (color-coded by redshift) with black crosses. The top and bottom panels on the right, respectively, show the normalized 1D spectra of the sources mapped to the top and bottom regions of the SOM. From this, it can be clearly seen how the SOM is effectively grouping similar galaxies closer together.

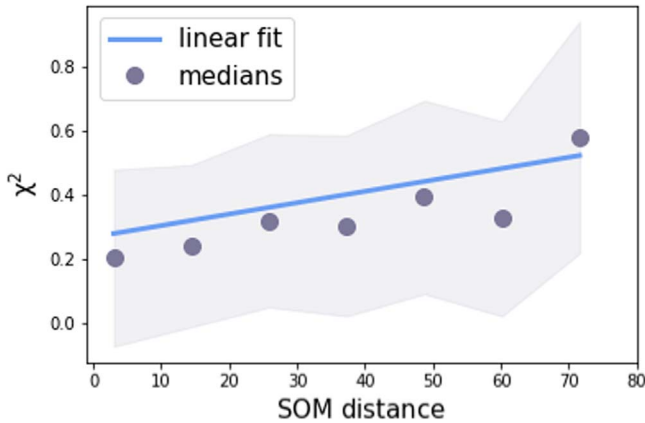


Figure 9. χ^2 difference between VVDS spectra vs. their distance on the SOM. We mapped multiband photometric SEDs of all sources with high-quality spectra from the VVDS to our trained SOM and measured distances of 2000 pairs of randomly selected galaxies with $dz = 1$. The blue solid line is the linear fit to the data. The gray circles and the shaded region represent the running median and 1σ dispersion, respectively.

scatter value for a 120×80 grid is ~ 0.02 . However, the larger the SOM grid is, trained by a limited number of training data, the more interpolations are forced by the SOM. In the above-mentioned grid of 120×80 , this leads to 262 cells ($\sim 3\%$) with no objects mapped to them and a minimum of $\sim 10,000$ total spectra requirement to fill each cell with at least one spectrum. The real LSST+*WFIRST* sample would have orders of magnitude more galaxies compared to CANDELS and therefore all of the cells will have sufficient data for the tomographic bins to have the cosmology-required accuracy. However, making the SOM too large would require even more spectroscopic observations for calibration, which is unnecessary given that the requirement is already met with a smaller SOM.

5.1. The C3R2 Survey

The C3R2 survey (Masters et al. 2017) is an ongoing spectroscopic effort to calibrate redshifts to the accuracy needed by weak lensing at the *Euclid* depth by comprehensively mapping the empirical galaxy color–redshift relation. This is a good sample to at least partially fill the voids of the spectroscopy on the *WFIRST* SOM as well. The C3R2 survey was initiated in 2015 with 10 allocated nights of Keck/DEIMOS observations (PI D. Stern) and five allocated nights of Keck/DEIMOS, MOSFIRE, and LRIS (PI J. Cohen), followed by more nights on Keck, the VLT, and the GTC later. Masters et al. (2017) released the first season of observations, including 1283 high-quality redshifts, and will release the second round of data in 2018 (Masters et al. 2019).

Unlike *Euclid*, which will use a broad optical (*riz*) filter to detect its weak-lensing shear sample, the *WFIRST* lensing sample will be derived from deep *H*-band observations, leading to a sample that includes optically faint sources well beyond the *Euclid* depth of $r_{\text{iz}} < 25$. Given the depth achieved by C3R2, in the following sections we explore the extra spectroscopy needed to calibrate sources fainter than what would be studied by *Euclid*.

5.2. How Different are Faint and Bright Galaxies?

The key question is whether the fainter galaxies comprising the *WFIRST* lensing sample have different SEDs or if they lie in the same color space as defined by *Euclid*. If the latter is the case, then the spectra acquired by the C3R2 to calibrate *Euclid* would be sufficient to address the *WFIRST* redshift calibration requirement as well. To first-order approximation, *Euclid*’s sample will contain galaxies with $r_{\text{iz}} < 25$. To answer this question, we divide the sample into bright ($r_{\text{iz}} < 25$) and fainter ($r_{\text{iz}} > 25$) subsamples. In Figure 11, we map the two

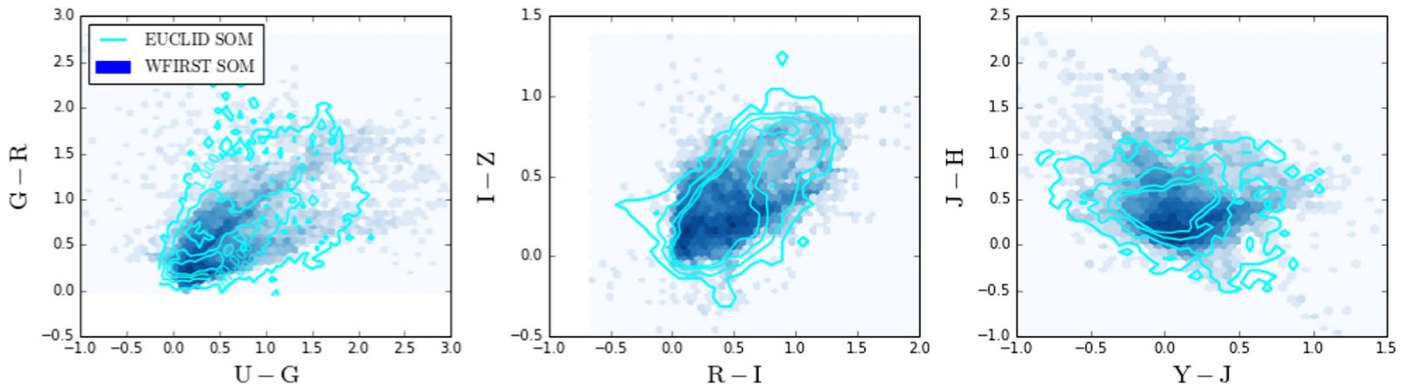


Figure 10. Comparison of the color distribution of the SOM cells trained with the *WFIRST* lensing sample (blue density) and *Euclid* sample (aqua contours, with [1, 10, 30, 100] contour levels). The *Euclid* sample is taken from the combination of the COSMOS, SXDS, and VVDS surveys covering a total of 3.8 deg^2 , 19 times larger than the area covered by the five CANDELS fields.

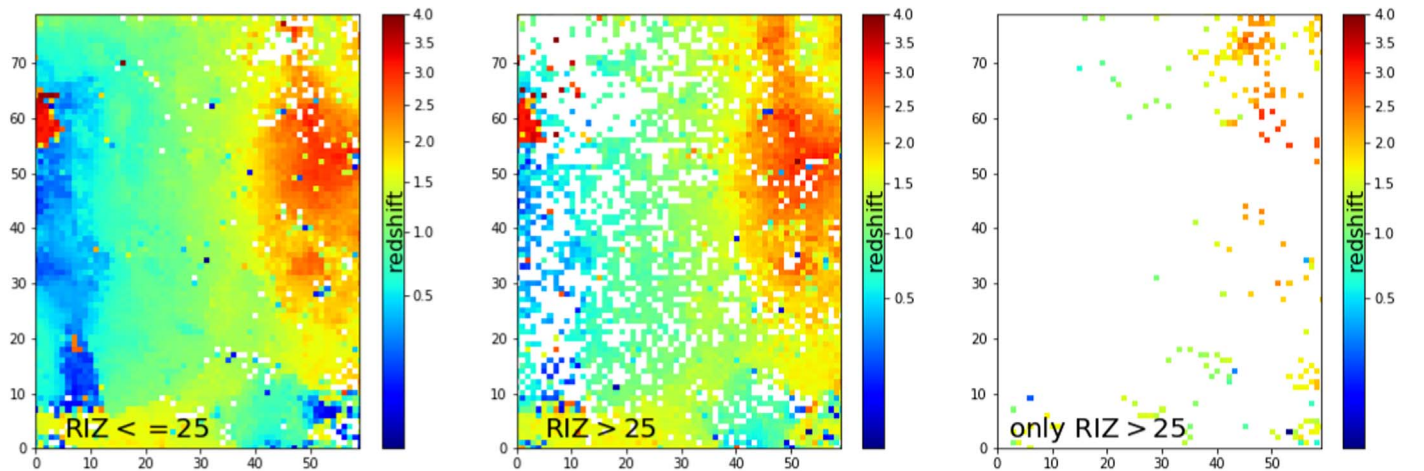


Figure 11. Bright ($\text{riz} < 25$; *Euclid* depth) and faint ($\text{riz} > 25$) galaxies in the *WFIRST* lensing sample are mapped to the SOM, color-coded by median redshifts (shown in the left and middle panels). More than 95% of the SOM cells contain at least one bright galaxy, $\sim 71\%$ of the SOM cells contain at least one faint object, and only $\sim 4\%$ of cells contain only faint galaxies (right panel).

subsamples onto the SOM separately (left panel: bright subsample, middle panel: faint subsample) and also identify cells that only contain faint galaxies (right panel). Approximately 95% of the SOM cells are covered by the bright sample, showing that fainter galaxies will not necessarily have different SEDs. While in the *WFIRST* lensing sample $\sim 26\%$ of objects have $\text{riz} > 25$, they spread over $\sim 71\%$ of the SOM, most of which also have bright galaxies mapped to them. The majority (91%) of the faint sample ($\text{riz} > 25$) live in color cells that are also occupied by brighter *Euclid*-depth sources ($\text{riz} < 25$), and only a small fraction of galaxies live in cells with no brighter counterpart ($\sim 4\%$ of the SOM cells, $\sim 2\%$ of the objects).

We explore the cells with no bright object and compare them to a neighboring cell. Figure 12 shows an example cell [8, 25] in the SOM with only faint objects mapped to it in comparison with its neighboring cell [7, 25] containing both faint and bright objects. As expected from the SOM (preserving topology), the SEDs of the two cells are very similar, and the difference in colors (~ 0.05) is negligible. Note that, as the SOM is trained with colors and the exact magnitudes of each cell are not fixed, we manually assigned F184W magnitudes to the two SOM SEDs in Figure 12 to plot the SEDs. The absence of bright objects in cells like these can then simply be due to the small area coverage of CANDELS fields, rather than new or different types of galaxies.

To examine this further, we compare the bright, faint, and the faint with no similar bright galaxies samples across the same redshift range ($1.8 < z < 2.2$) on color-color plots in Figure 13. The redshift range is fixed to eliminate the effect of distance on brightness. As can be seen in these color-color plots, fainter points with no similar bright SEDs (red data points) have much larger photometric errors compared to the bright objects (blue data points). As demonstrated before, faint objects do not occupy different portions of color space and have similar colors to the rest of the objects within their errors. Note that, in the lower right panel of Figure 13, where the *Spitzer*/IRAC Ch1 – Ch2 color is plotted on the x axis, there is a distinct class of bright objects with no fainter counterparts (as per Stern et al. 2005, presumably active galactic nuclei), while all fainter objects do have a neighboring bright object with similar colors. In short, fainter objects in our *WFIRST* lensing sample live in the same color space defined by brighter objects.

5.3. Redshift Accuracy as a Function of Brightness

We have demonstrated that the majority of faint and bright galaxies live in the same color space, with similar SEDs. However, to be able to calibrate the *WFIRST* sample with the C3R2 sample, it is important to test for redshift accuracy as a function of brightness. Extra spectroscopy per cell is needed if

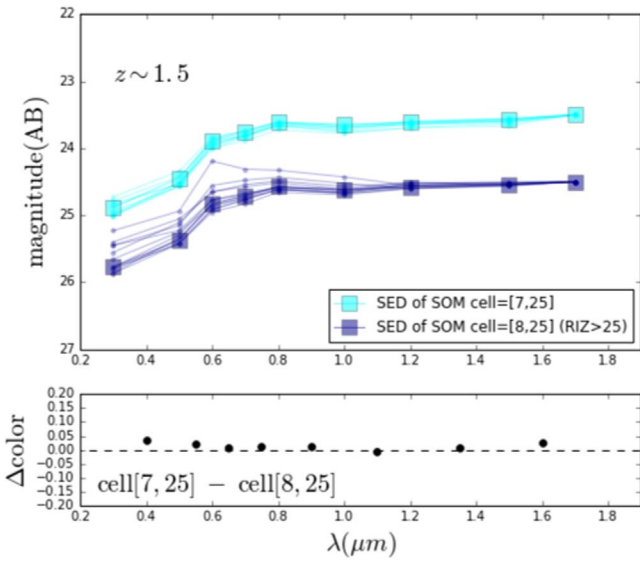


Figure 12. SED of a SOM cell [8, 25] containing only faint ($\text{riz} > 25$) objects, shown with dark blue squares and the SED of objects mapped to this cell presented as dark blue dots and solid lines, is compared to its neighboring cell [7, 25] containing both bright and faint objects, shown with aqua squares and the SED of objects mapped to this cell presented as aqua dots and solid lines. The F184W magnitude of the SOM cell SEDs are assigned manually to demonstrate bright vs. faint, and the rest are from the colors of the cell. The bottom panel shows the color differences between the two cells, with the x axis being the average wavelength between the two filters of each color.

the redshift scatter gets much larger as we include more faint galaxies to the SOM cell.

In the top panels of Figure 14, we test for redshift accuracy as a function of brightness by simulating galaxies with similar SEDs at different brightnesses and measuring their redshifts. In this test, we chose an SOM cell with $\text{spec-}z = 1.28$ having both faint and bright galaxies (four with $\text{riz} < 25$ and three fainter than this threshold) from the lensing sample mapped to it. For each of these galaxies, we generated 1000 similar SEDs within their uncertainties at each band. We measured the redshift of each simulated SED using the SED-fitting code LEPHARE (Amouts et al. 1999; Ilbert et al. 2006) as well as using the SOM (i.e., SOMz). The middle panel of Figure 14 shows the distribution of the photometric redshifts for simulated galaxies in different bins of riz magnitude. This panel shows that moving to fainter magnitudes, the fraction of wrong redshift assignments gets significantly larger, as seen from the second peak in the distribution. The blue vertical lines on the violin plots, representing the 1σ value of the redshift distributions, are larger than allowed for redshift calibration in each cell. Therefore, SED-fitting codes will not be able to provide the redshift uncertainty required for the fainter subsamples. However, as we show in the right panel of Figure 14, if the color–redshift relation from the SOM is used to measure the redshifts (SOMz), the scatter significantly decreases. SOMz is measured by mapping the simulated galaxies to the SOM. As can be seen from this figure, there is no bias in the median redshifts with brightness when using the SOM method and the dispersions are also of the order (~ 0.04) allowed by the SOM calibration technique.

We note that in the SOMz method here, the mapping of color to redshift is generated from the median CANDELS photo- z of galaxies in each cell. Ideally, once the SOM is covered with spectroscopic redshifts, the mapping would be more accurate

and less dependent on the prior use of SED-fitting codes as is the case here. To assure that the improvement we see by using the SOMz method is not fully based on the prior use of the same galaxies in training the SOM and making the color–redshift relation, we extend our test (bottom panels of Figure 14) to galaxies from the COSMOS catalog (Laigle et al. 2016). We used the nine closest photometric bands to generate and map the colors to our SOM. We measured the photo- z and SOMz of a thousand realizations of the 21 galaxies, which map to that same cell on the SOM. While these galaxies are all relatively bright ($\text{riz} < 25$) and there are slight discrepancies in the photometries used (e.g., Ultravista near-IR photometries in COSMOS versus *HST* in CANDELS), the improvement from traditional SED fitting is still evident.

5.4. Spectroscopy Recommendation

As shown in the previous sections, spectroscopic sampling by the C3R2 survey for the *Euclid* mission is sufficient for filling more than 90% of the color space with at least one spectra. To fill out the remainder of the *WFIRST* SOM, 200 new spectroscopic redshifts are needed to fill the color space with at least one spectra at each cell. This corresponds to the 4% of the cells that have only faint objects associated with them. With 1200 spectroscopy, cells with larger dispersions (25% of cells where redshift scatter > 0.05) will have two spectroscopic redshifts mapped to them. This would be helpful in cells where most of their associated galaxies are faint. Therefore, we recommend ~ 200 –1200 extra spectra to fully calibrate the photometric redshifts of the LSST+*WFIRST* lensing sample.

In addition to weak-lensing calibration, from the galaxy evolution point of view, it is absolutely important to obtain spectroscopic observations of these faint systems, which have not been spectroscopically observed before. This will enable the comparison of their more detailed physics to brighter galaxies with similar broadband SEDs (colors).

The number of recommended spectra needed to calibrate the LSST+*WFIRST* color–redshift relation is not large. However, spectroscopy of these faint targets would not be easy. Most of the voids in previous spectroscopic observations that we found by using the SOM are likely due to the biased selection techniques. Another possibility, however, can be due to unsuccessful spectroscopic observations, i.e., insufficient observing time or wrong telescope/instrument chosen for the observations. Most of the cells with no spectroscopic coverage are in the $z \sim 1$ –2 regime, once dubbed the redshift desert, due to historical difficulties. Powerful instruments on ground and space-based telescopes such as the Keck twin telescopes, the future Thirty Meter Telescope, and the *James Webb Space Telescope* can now/will explore this redshift regime and to fainter depths.

6. Summary

In this paper, we studied the redshift calibration requirements for *WFIRST* HLS weak-lensing analysis to meet the desired stage IV dark energy accuracy. We adopted the methodology introduced by M15, which calibrates the color–redshift relation using the SOMs. We imitated the LSST+*WFIRST* lensing sample using optical and near-IR data from the five CANDELS fields and trained an SOM with successive colors of galaxies in

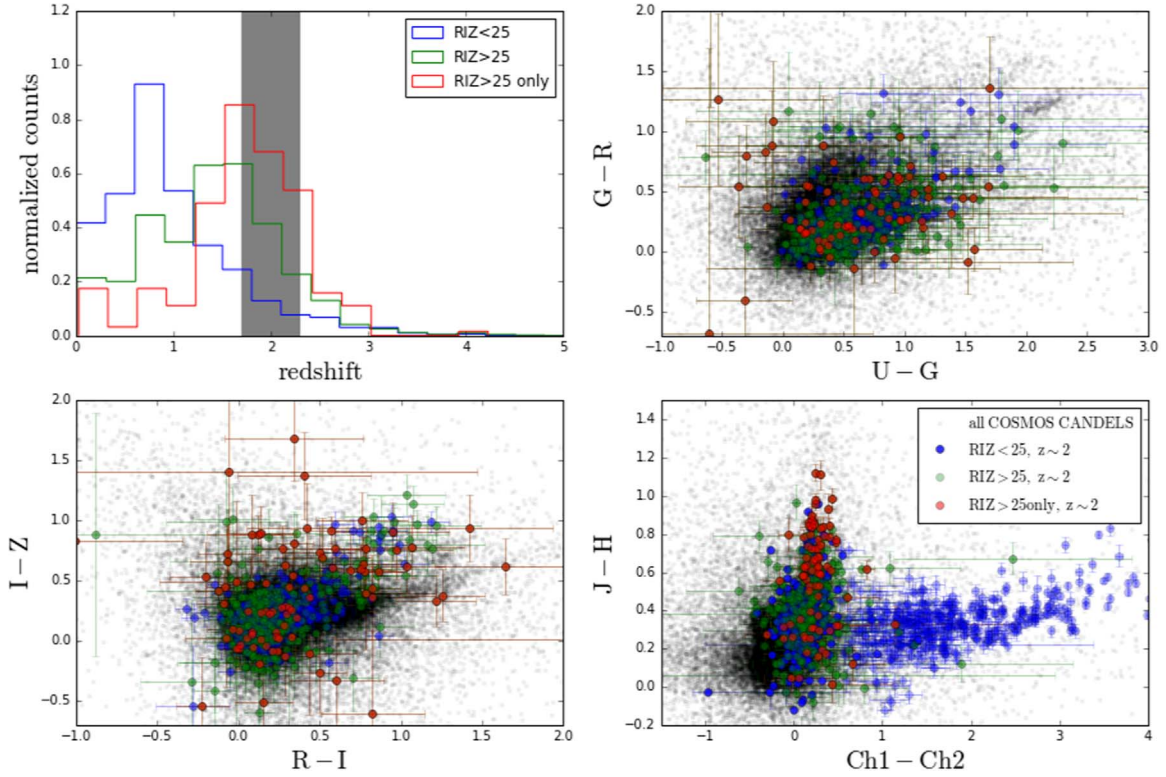


Figure 13. Comparison of the three subsamples in our *WFIRST* lensing catalog: bright ($\text{riz} < 25$; blue), faint ($\text{riz} > 25$; green), and only faint ($\text{riz} > 25$, and no bright SED in the object’s SOM cell; red) galaxies in the same redshift range in the color–color plots. Top left: the normalized distribution of subsamples is shown and the selected redshift range ($1.8 < z < 2.2$) for comparison in the other three subplots is shown as the shaded gray region. The $G - R$ vs. $U - G$ (top right), $I - Z$ vs. $R - I$ (bottom left), and $J - H$ vs. *Spitzer*/*IRAC* Ch1 – Ch2 (bottom right) plots show the distribution of all COSMOS CANDELS galaxies in black and overplotted are the three subsamples.

the LSST+*WFIRST* filter set. The smoothness of the redshift distribution on an SOM trained by colors illustrates the color–redshift relation and makes the SOM an optimal source for spectroscopic target selection.

Based on Monte Carlo simulations in M15, and given the estimated average redshift uncertainties in our SOM cells, a tomographic bin containing ~ 200 SOM cells would be sufficient to reach $\Delta\langle z \rangle / (1 + \langle z \rangle) < 0.002$. For the technique to be efficient, most SOM cells need to have at least one spectroscopic object mapped to them for calibration. This is equivalent to ~ 5000 total spectroscopic redshifts to calibrate the *WFIRST* SOM. However, in addition to the already existing spectroscopic observations in the CANDELS fields (covering 57% of the SOM cells), the C3R2 survey is filling the color space of *Euclid* galaxies with spectroscopic observations.

We showed that $\sim 26\%$ of the *WFIRST* lensing sample consists of sources fainter than the *Euclid* depth in the optical, 91% of which live in color cells also occupied by brighter (*Euclid*-depth) sources. We demonstrated the similarity between the fainter and brighter subsamples in the same cells as well as the feasibility of measuring the redshifts of fainter objects to the accuracy needed using the SOM color–redshift relation. Because the $\sim 4\%$ of cells that have only fainter objects associated with them might be due to the small sample size in the CANDELS fields as well as larger photometric errors in the fainter sample, we recommend extra spectroscopy for these cells to calibrate the color–redshift relation on *WFIRST* SOM thoroughly. We recommend ~ 200 – 1200 new spectra, which will cover the cells with only faint objects as

well as those with large redshift dispersions. It is crucial to note that having most of the calibration already in place with the C3R2 *Euclid* effort does not imply similarity between *WFIRST* weak-lensing cosmology and the less deep surveys, as the lensing sample size will increase significantly (see Figure 15).

In our analysis, we have used an interpolation technique to estimate the photometry in the LSST+*WFIRST* filter sets based on the available photometry in different filter sets. The method is tested extensively and is the easiest logical way to reproduce statistically correct distributions of photometries and colors of galaxies. Future works using improved techniques will test for the robustness of the interpolation on an object by object basis.

One weakness of this work is the small sample size used for training the SOM. As discussed in the paper, CANDELS data were the best available option, due to the comparable depth of the observations to those expected from the *WFIRST* HLS. CANDELS observations being done in five well-separated fields in the sky should mitigate the effect of cosmic variance. We found the color space of our *WFIRST* lensing sample to be representative of a sample trained for *Euclid* (19 times larger in area), which suggests that the effect of cosmic variance on the SOM calibration technique should be minimal. However, very large area simulations are needed to enable a more quantitative investigation of this effect on the findings presented in this work (P. Capak et al. 2019, in preparation). We will revisit our forecasts once *WFIRST* observations are available and retrain an SOM with actual observations for photometric redshift calibration as well as the selection of weak-lensing tomographic bins.

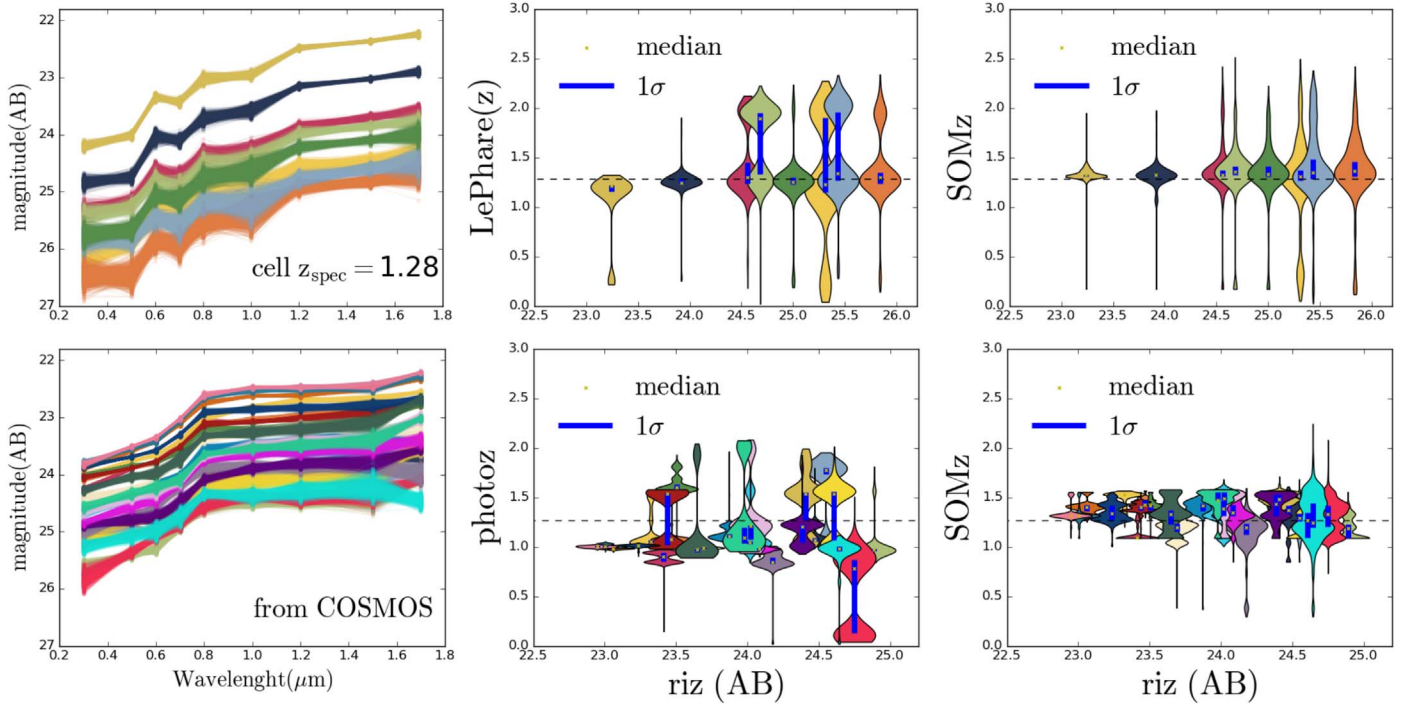


Figure 14. Top left: SED of seven galaxies with different brightnesses but similar SED shapes mapped to a cell at spec- $z = 1.28$ are shown in different colors. From each SED, we extract 1000 (Monte Carlo) realizations within the observed uncertainties (as seen from the thickness of each SED). The larger spread of simulated SEDs in the fainter galaxies is due to the larger photometric errors associated with them. Top middle: photometric redshifts of each group of galaxies measured by the LePhare code colored with the colors of the SEDs in the left panel. The blue vertical lines show a 1σ spread around the median. The fraction of wrong redshift estimates gets larger as one moves to fainter magnitudes, due to larger uncertainties, as can be seen from the second peaks mostly at higher redshifts. Top right: same as the middle panel with SOMz for each group of galaxies measured by mapping the simulated SEDs to the SOM. Redshift estimates get significantly better compared to the middle panel, with a much smaller fraction of outliers. Bottom panels are the same as the top panels, with 21 COSMOS galaxies mapped to the same cell of our trained LSST+*WFIRST* SOM.

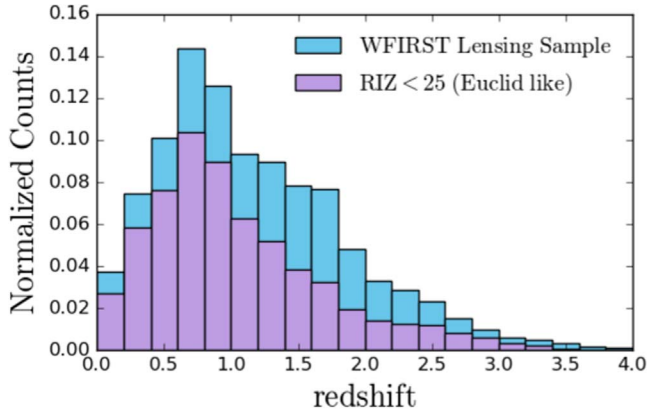


Figure 15. *WFIRST* will significantly increase the lensing sample size. The normalized redshift distribution of galaxies in the LSST+*WFIRST* analog sample is shown as the blue histogram, and the fraction of galaxies with $r_{\text{IZ}} < 25$ as would be found by *Euclid* is overplotted in purple.

We wish to thank the anonymous referee for constructive comments that significantly improved this work. S.H. is grateful to Chris Hirata for constructive comments and for providing the lensing sample selection criteria. S.H. also thanks Rebecca Larson for carefully reading the manuscript. This work used SOMPY, a Python package for self-organizing maps (main contributors: Vahid Moosavi, @sevamoo; Sebastian Packmann, @sebastiandev; Iván Vallás @ivallesp). This research made extensive use of data from the CANDELS survey. Parts of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology,

under a contract with the National Aeronautics and Space Administration.

ORCID iDs

Shoubaneh Hemmati <https://orcid.org/0000-0003-2226-5395>
 Peter Capak <https://orcid.org/0000-0003-3578-6843>
 Daniel Masters <https://orcid.org/0000-0001-5382-6138>
 Iary Davidzon <https://orcid.org/0000-0002-2951-7519>
 Olivier Doré <https://orcid.org/0000-0001-7432-2932>
 Jason Rhodes <https://orcid.org/0000-0002-4485-8549>
 Daniel Stern <https://orcid.org/0000-0003-2686-9241>

References

- Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv:astro-ph/0609591
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
- Aubourg, É., Bailey, S., Bautista, J. E., et al. 2015, *PhRvD*, 92, 123516
- Benítez, N., Gaztañaga, E., Miquel, R., et al. 2009, *ApJ*, 691, 241
- Blandford, R. D., & Narayan, R. 1992, *ARA&A*, 30, 311
- Blandford, R. D., Saust, A. B., Brainerd, T. G., & Villumsen, J. V. 1991, *MNRAS*, 251, 600
- Bordoloi, R., Lilly, S. J., & Amara, A. 2010, *MNRAS*, 406, 881
- Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, *ApJS*, 200, 13
- Carrasco Kind, M., & Brunner, R. J. 2013, *MNRAS*, 432, 1483
- Cunha, C. E., Huterer, D., Busha, M. T., & Wechsler, R. H. 2012, *MNRAS*, 423, 909
- Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, *ApJ*, 775, 93
- Doré, O., Hirata, C., Wang, Y., et al. 2018, arXiv:1804.03628
- Faisst, A. L., Capak, P., Hsieh, B. C., et al. 2016, *ApJ*, 821, 122
- Furusawa, H., Kosugi, G., Akiyama, M., et al. 2008, *ApJS*, 176, 1
- Galamez, A., Grazian, A., Fontana, A., et al. 2013, *ApJS*, 206, 10
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35

- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, [ApJS](#), **207**, 24
- Hu, W. 1999, [ApJL](#), **522**, L21
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, [A&A](#), **457**, 841
- Kitching, T. D., Taylor, A. N., & Heavens, A. F. 2008, [MNRAS](#), **389**, 173
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, [ApJS](#), **197**, 36
- Kohonen, T. 1982, *Biological Cybernetics*, **43**, 59
- Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, [ApJS](#), **218**, 15
- Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, [PASP](#), **119**, 1325
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, [ApJS](#), **224**, 24
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, [arXiv:1110.3193](#)
- Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, [A&A](#), **559**, A14
- Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, [A&A](#), **439**, 845
- Masters, D., Capak, P., Stern, D., et al. 2015, [ApJ](#), **813**, 53
- Masters, D. C., Stern, D., Cohen, J. G., et al. 2017, [ApJ](#), **841**, 111
- Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, [arXiv:1904.06394](#)
- Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, [ApJS](#), **228**, 7
- Newman, J. A. 2008, [ApJ](#), **684**, 88
- Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, [Aph](#), **63**, 81
- Newman, J. A., & Davis, M. 2002, [ApJ](#), **564**, 567
- Oke, J. B., & Gunn, J. E. 1983, [ApJ](#), **266**, 713
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, [ApJ](#), **517**, 565
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, [AJ](#), **116**, 1009
- Sánchez, C., & Bernstein, G. M. 2019, [MNRAS](#), **483**, 2801
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, [ApJS](#), **172**, 1
- Spergel, D. N. 2015, [Sci](#), **347**, 1100
- Stefanon, M., Yan, H., Mobasher, B., et al. 2017, [ApJS](#), **229**, 32
- Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, [ApJ](#), **631**, 163
- Zhan, H., & Knox, L. 2006, [ApJ](#), **644**, 663